SESx : a program for analytic computation and exact triangulation of solvent-excluded surfaces

Lincong Wang

Abstract

The solvent-excluded surface (SES) of a molecule is an inherent geometrical property of its structure relevant to its solvation and stability in aqueous solution, and its functions. This document describes a robust program, SESx, for analytic SES computation and exact triangulation. The unprecedented accuracy and robustness of SESx make it possible to analyze and to visualize in great detail the SESs of any molecules in general and biomolecules in particular. Compared with other SES computation/triangulation program only SESx is capable of computing, triangulating and displaying, *individually*, each of the 3 types of SES.

1 The usage

sesx 1bl8FH.pdb > 1bl8ses.log

MOUSE: LEFT for rotation, RIGHT for translation and MIDDLE(WHEEL) for zooming

2 The terminal output

We use the crystal structures of a potassium channel (IBL8FH.pdb) and CRABP-II (2FR3FH.pdb) as examples to demonstrate SESx.

2.1 Some basic information about the structure extracted from the PDB file

```
HEADER MEMBRANE PROTEIN 23-JUL-98 1BL8
TITLE POTASSIUM CHANNEL (KCSA) FROM STREPTOMYCES LIVIDANS
REMARK 2 RESOLUTION. 3.20 ANGSTROMS.
HETNAM K POTASSIUM ION
———-
1bl8FH Chains:
( B 1412) U+2204_atom = 57, 4.04% {Missing 57 atoms, i.e. about 4.04% atoms are missing}
( D 1412) U+2204_atom = 57, 4.04%
( C 1412) U+2204_atom = 57, 4.04%
( A 1412) U+2204_atom = 57, 4.04%
**********
Atom_Probe_CPUtime(s)_clockTime(s): 5648   6720   11.042   1.951
```

The last row lists respectively the total numbers of atoms in 1BL8 and the total numbers of probes computed by SESx, the CPU time and the clock time in second for SES computation. In general clock time (with multiple threads) is about 5-fold shorter than the CPU time.

3 Main window and Keys

The default main window: SES with all three components (SAS, toroidal and probe SES) shown in gray color.

In SESx each key or key-combination (i.e. control + key-l, shift + key-l etc) just toggles between show and hide, i.e. *On* or *Off*. Most of common keys are listed below. Other keys will be described late when they are required.

key-e: entire SES
key-v: colors the 3 SES types with SAS in pink, torus in yellow/spindle-torus in magenta and probe SES in blue
key-d: colors the SASs by the value of atomic partial charge with blue for the most negative and red for the most positive
key-a: shows SASs
key-p: protein with backbone as default, then key-h, key-m show respectively heavy and all the atoms
key-s: protein space-filled model
key-c: colors the protein atoms with C, N, O, H, S depicted respectively in green, blue, red, gray and yellow
Control + *key-m*: shows metals as sphere
key-y: shows (chemical) compounds
key-j: shows compounds in stick-and-ball or sphere
key-z: shows compound-protein interface

4 Two SESs by SESx

In the following we select two crystal structures, 1BL8 and 2FR3, to demonstrate some of the unique features of SESx.

4.1 The SES of 1BL8 by SESx

With a gray-colored SES whether the gray-colored space-filled model is On or Off makes no difference in visualization. However, if the protein atoms are colored, then the differences between SAS triangulation and atom triangulation become evident. With SASs Off but protein atoms On and colored, all the solvent-accessible atoms (SA-atoms) are shown as a spherical polygon. However since the polygons shown above are not defined analytically by their sides light will possibly leak.

With the SASs colored by charge value (*key-d*) it is obvious that the aqueous solvent accessible surfaces (i.e. solvent-protein interface) of 1BL8, i.e. both the extracellular side and intracellular (cytoplasmic) side of this membrane protein, have negative average partial charge. In fact, having a negative average partial charge is a salient feature of almost all the soluble proteins.

On the other hand, the protein atoms on the lipid-interface are mainly carbons with slightly negative charges and protons with slightly positive charges. Protein atoms with either large negative or large positive charges, if exist, almost always have very small SAS area: a characteristic shared by all the integral membrane proteins. In addition, a salient geometric feature of the lipid-protein interface is that its SES is less rugged than the SES of a purely solvent-accessible interface.

4.1.1 The potassium channel viewing from the exterior

Let's start with the extracellular side. The near perfect C4 symmetry as illustrated by the identical locations, shapes, colors, areas of each of the quadruples of SASs suggests that the tetramer is constructed using a monomer via rotation and translation.

Control + *key-m* shows the potassium atom (yellow-colored) with a 2.03Å (covalent) radius. The entrance to the channel is composed of mainly atoms (oxygens) with large negative partial charges, as shown by *key-p* (brings back the space-filled model) and *key-a* (hides the SASs). The exit to the channel (on the intracellular side) is formed by 4 Qs with their carbonyl oxygens exposed.

4.1.2 The potassium channel viewing from the interior

Now let's go back to the SES surface, hide the metals and zoom into the interior of the channel. In the present implementation, it is somewhat awkward to show the interior. It requires some patience. When getting into the interior from the extracellular side we first encounter an empty space, then a large cavity. Please remember, here empty space means the space where the protein atoms are packed so tightly that no cavities larger than a probe exist. Getting inside the cavity, we see a hole (a circle) < 1.4Å radius that leads to the exit. *key-v* shows that the hole is inside a probe SES polygon. Most of current analytic SES program (e.g. MSMS) ignore such an interior circle.

Let's start with the exit and get into the interior there. It has an interior-region formed by two arcs each from one of a pair of intersecting probes. However, the two arcs are not perfectly symmetrical. It suggests that C4-symmetry has not been constructed exactly. Even minor differences among the monomers in the assembled tetramer are discernible in `SESx` : demonstrating the accuracy of `SESx`.

Now let's get into the interior from the side of the channel. Let's first display the 3 potassium atoms (*Control* + *key-m*) and color the SASs by charge value (*key-d*). The 3 potassium (3 Ks) identify the location of the channel. Hiding the Ks, viewing from the interior the entrance looks like a large lamp hanging from the ceiling. There are empty space between the first K and the other 2 Ks. The channel where the other 2 Ks occupy looks like a mushroom head while the bottom K atom seems much larger than the size of the SES cavity. It is because the probe radius is only 1.4Å while the radius of K is 2.03Å.

One structure feature that could be easily identified via `SESx` is that there are several large cavities surrounding the completely-filled (or the narrowest) portion of the channel. Since some movements must occur for potassium to pass, the existence of such cavities make such movements feasible.

4.1.3 The PPI interfaces

Now let's take a look at the protein-protein interaction (PPI) interface.

First turn off SES color and use *key-i* to show the protein interface atoms in color. The probes all of whose three constituent atoms are PPI interface atoms could be shown using *Shift* + *key-i*. We call them *ppi-probes* and are shown as spheres with a small radius. The locations of these ppi-probes trace out PPI interface in 3D.

Let's walk through one PPI interface (a cleft) starting at the extracellular entrance. For clarity let's hide the ppi probes. As illustrated here the SES of a PPI cleft is, in general, more complicated geometrically (with holes, pockets and passes) than the SES of a purely solvent-accessible interface (i.e. solvent-protein interface). The existence of such complicated geometrical features shows that a typical PPI interface is less tightly packed than that of a purely solvent-accessible interface. In addition the interior of a PPI interface is also generally more complicated than a monomer interior.

With only the SASs for the interface atoms colored, viewing from the interior it is obvious that this potassium channel is formed only by interface atoms.

4.2 The SES of 2FR3 by `SESx`

In this example we use the crystal structure, 2FR3, to demonstrate that the analytically-computed and exactly-triangulated SES from `SESx` in combination with other geometrical and physical properties could be used to quan-

tify a ligand-protein interface, i.e. ligand binding site.

2FR3 is the crystal structure of human cellular retinoic acid binding protein type II (CRABP-II) in complex with all-trans-retinoic acid (RA). The solution structures and the RA-protein interface of human CRAPB-II are the main focus of my Ph.D. research. In general as for any soluble proteins, the SASs of the surface atoms of CRABP-II with large negative charges have also large SAS areas while those with positive charges or small negative charges (i.e. carbon atoms) have small SAS areas.

key-y shows the compound (RA) without protons in stick model while *key-j* shows RA in space-filled model. *Control+key-i* shows the SASs (colored by charge value) of the binding site atoms. *key-z* shows the binding site residues while *Shift+key-l* labels them. In each residue the atoms of < 3.0Å away from any ligand atom are shown as a small ball. *key-u* shows them with their VDW radii. The SASs with negative partial charge and large SAS area are located at the entrance while the SASs with positive charge and (likely) large SAS area are at the bottom. In particular the three atoms with (likely) large SAS area and positive charge form H-bonds with the carboxyl group of RA. Except for the two ends, the binding site consists mainly of carbons and protons with close to zero partial charges and small SAS areas, i.e. a hydrophobic region with relatively flat surface. When viewing from the interior, one carboxyl oxygen protrudes into the SES because it forms an H-bond with Y134-OH. In addition there exist some extra space in the bottom of binding site, and near the bottom, there are relatively large cavities. They suggest that the bottom of the binding site is somewhat flexible. In addition, there is an additional channel leading from the bottom to the surface, again suggesting the flexibility of the RA binding site. One side of the binding site (like a channel).